B

**1000CST473112401**

Pages: 2



Reg No.:_____ Name:_____

**APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY**

B.Tech Degree S7 (R, S) Examination November 2024 (2019 Scheme)

**Course Code: CST473**

**Course Name: NATURAL LANGUAGE PROCESSING**

**Max. Marks: 100** **Duration: 3 Hours**

## PART A

***Answer all questions, each carries 3 marks.*** Marks

1 Explain the key tasks and applications of NLP. Provide examples of how NLP is applied in real-world scenarios. (3)

2 Differentiate between heuristics-based NLP and machine learning-based NLP. Discuss the strengths and limitations of each approach. (3)

3 Explain the steps involved in the NLP system pipeline. How do data acquisition, pre-processing, and feature engineering contribute to building effective NLP models? (3)

4 Consider two documents in a corpus: (3)
**Document 1**: "AI in healthcare is transformative."
**Document 2**: "Healthcare and AI are creating new solutions."
Calculate the TF-IDF score for the word **"AI"** in these two documents.

5 How does Naïve Bayes perform in sentiment classification, and what is its training process? (3)

6 Discuss the challenges of ambiguity in NER and explain how NER is approached as a sequence labelling task. (3)

7 How do supervised learning approaches differ from lightly supervised approaches in relation analysis? (3)

8 What is an inverted index in information retrieval? Discuss its role in term weighting and document scoring, and how it improves retrieval efficiency. (3)

9 Describe the concept of Mean Reciprocal Rank and explain how it is used to evaluate information retrieval systems. (3)

10 Provide an algorithm for performing word alignment in Machine Translation. How does it contribute to improving translation quality? (3)

## PART B

***Answer any one full question from each module, each carries 14 marks.***

### Module I

11 a) Explain the Naïve Bayes classifier in the context of NLP. How does it work, and what are its advantages and limitations? (7)

b) Describe the logistic regression model for NLP tasks. How does it handle binary and multi-class classification problems? (7)

**OR**

12 a) What are the main challenges faced in Natural Language Processing, and how can they be addressed? (7)

b) Explain the different building blocks of a language and their roles in NLP. (7)

## Module II

13 a) What is Doc2Vec, and how does it extend the concept of word embeddings to document-level representation? Discuss its application in NLP tasks. (7)

b) What is the significance of feature engineering in NLP? Provide examples of different features used and explain how they impact the performance of NLP models. (7)

**OR**

14 a) Compare the advantages and limitations of traditional text representation methods (such as One-Hot Encoding and Bag of Words) with advanced techniques like Word Embeddings and Doc2Vec. How do these methods influence model training and accuracy? (14)

## Module III

15 a) Discuss the role of Logistic Regression in text classification. How does it perform in both binary and multi-class classification problems, and what makes it suitable for text data? (7)

b) What is Information Extraction (IE)? Explain the general pipeline for building an IE system and provide examples of its applications in real-world scenarios. (7)

**OR**

16 a) Discuss the differences between Naïve Bayes, Logistic Regression, and SVM in the context of text classification. Evaluate their strengths and limitations across various text classification tasks. (14)

## Module IV

17 a) What is information retrieval in NLP? Explain the process of term weighting and document scoring, and discuss their importance in improving retrieval performance. (14)

**OR**

18 a) Explain the evaluation process for information retrieval systems. How are metrics like precision, recall, and Mean Average Precision (MAP) used to measure retrieval system performance? (14)

## Module V

19 a) What are question-answering systems? Explain the different stages involved in factoid question answering, including question processing, passage retrieval, and answer processing. (14)

**OR**

20 a) Explain the concept of classical machine translation. How do direct translation and transfer-based approaches differ, and what are their respective strengths and weaknesses? (14)

****